Alveolar macrophage transcriptomic profiling in COPD shows major lipid metabolism changes

Wataru Fujii, Theodore S. Kapellos, Kevin Baßler, Kristian Händler, Lisa Holsten, Rainer Knoll, Stefanie Warnat-Herresthal, Marie Oestreich, Emily R. Hinkley, Jan Hasenauer[6], Carmen Pizarro, Christoph Thiele, Anna C. Aschenbrenner, Thomas Ulas, Dirk Skowasch, Joachim L. Schultze

ONLINE DATA SUPPLEMENT

METHODS

Human specimens

Human studies were approved by the ethics committees of the University of Bonn and University hospital Bonn (local ethics vote 076/16). All patients provided written informed consent according to the Declaration of Helsinki before specimens were collected. They were recruited as part of an exploratory observational clinical cohort study consecutively over a period of 27 months from the Department of Pneumology. Patients were categorized as either control (see definition below), COPD, or other diseases (see also Fig. 1A). Patients with COPD were diagnosed and stratified according to the guidelines of the global initiative for chronic obstructive lung disease (GOLD) [1]. Eligible patients were aged 18 years or older and were either current, past or non-smokers (Table E1). Current smokers had smoked in the last 3 months, ex-smokers had not smoked in the last 3 months prior the bronchoscopy and never smokers had not smoked more than 100 cigarettes in their lifetime and did not smoke at the time of recruitment. Age-matched individuals suffering from chronic idiopathic cough, demonstrating an exquisitely sensitive cough reflex without underlying pathology [2], served as control donors. A diagnostic algorithm that considered medical history (including drugs, e.g. ACE inhibitor and indications of heart burn), physical examination, echocardiography, chest X-ray, lung function, including methacholine challenge, presentation at an HNO doctor, blood test (including exclusion of eosinophilia), FeNO (excluding >50 ppb), computer tomography of the chest and finally bronchoscopy [3] was worked up to enroll the control group. Exclusion criteria included hypoxemia despite oxygen supplementation (O_2 saturation<90%), hypercapnia, increased risk of bleeding, unstable cardiac disease, and COPD exacerbations within

the 4 weeks prior to recruitment. Patients with other pulmonary diseases (termed other (Fig. 1A)) were diagnosed as asthma, ACO, bronchiectasis, cancer, fibrosis, pneumonia and sarcoidosis (Table E2), but were excluded from further evaluation within this study.

Bronchoscopy procedure

Bronchoscopy was performed as a part of the diagnostic workup by two bronchoscopists through oral access and with light conscious sedation. All participants received supplemental oxygen by nasal cannula. After a general inspection, BAL was performed in the middle lobe or, if not accessible, the lingular lobe. Warmed saline (6 syringes of 20 ml each) was instilled into the airways to enable BALF recovery. After each instillation, the aliquot was immediately recovered by gentle hand suction into a syringe.

BALF processing

Human BALF was obtained from all patients included in the study (control, COPD, other) through bronchoscopy. Within this report, only BALF samples of the highest quality following quality criteria established previously for the processing of BALF [4], such as recovery rate higher than 30% and absence or minimal blood/mucus contamination, from control and COPD patients were used further. BALF specimens were washed with PBS, suspended with 0.02% EDTA-2Na and washed again for final re-suspension with 2% FCS/1 mM EDTA. $CD45^{+}Lin^{-}HLA\text{-}DR^{+}CD66^{-}Autofluorescence^{+}$ AMs were sorted using a FACS Aria III cell sorter (BD Biosciences, USA).

Cell counting and decision making

Total cell counts were determined with (1:5) Trypan Blue exclusion (Sigma-Aldrich) under an optical microscope. BALF cells (10ul) were diluted 1:10 in the Trypan Blue solution and counted in a Neubauer haemocytometer. Due to material limitations, we prioritized flow cytometry-based immunophenotyping of total $CD45^+$ cells. If sufficient material was available, we also sorted alveolar macrophages from the remaining cells for RNA sequencing and lipidomics. For some patients, the absolute numbers of alveolar macrophages were not sufficient for both approaches (transcriptomics and lipidomics) and we therefore proceeded to either RNA sequencing or lipidomics (Table E3).

Flow cytometry/FACS

Single-cell suspensions were stained with Live/Dead yellow fluorescent dye (ThermoFisher, USA) for 15 min at room temperature and were washed with PBS at 300xg for 5 min at 4^oC. They were then re-suspended in 100 ul PBS and blocked with 5 ul human FcR blocking reagent (Miltenyi, Germany) for 15 min on ice and were subsequently stained with the listed anti-human antibodies (Table E4) in buffer containing PBS, 2% FCS, 1 mM EDTA for 30 min on ice. After an initial interim analysis of the first 36 samples, we identified a non-annotated cell population, which we identified as $FcεRI^+$ mast cells. All further samples from control and COPD patients (n=23) were also interrogated for the presence of mast cells in the BALF. Cells were spinned at 300xg for 5 min at 4^oC and re-suspended in buffer containing PBS, 2% FCS, 1 mM EDTA for analysis. Data acquisition was performed on a FACS Aria III cell sorter

(BD Biosciences, USA). Analysis was performed with FlowJo v.10 software (Tree Star, USA).

Cytospin preparation

Cytospins were obtained by centrifuging $2x10^5$ cells in 200 ul PBS on microscope slides at 20% power for 5 min. Excess buffer was carefully discarded and slides were air dried for 3 min followed by fixation with 100% methanol for 5 min at 4^oC. The slides were subsequently washed with PBS and stained with 1:20 Giemsa solution (Sigma, USA) for 25 min at room temperature. A final rinsing step with H_2O and air drying before mounting followed. Cell morphology was examined by microscopic evaluation of stained cells using an Axio Lab A1 microscope (Zeiss, Germany).

RNA extraction and library preparation

Total RNA was isolated from human AMs with the miRNeasy Micro kit (Qiagen, Germany) according to the manufacturer's protocol and RNA concentration and integrity was determined using the High Sensitivity RNA assay on a TapeStation 4200 system (Agilent, USA). All samples had RIN>7 and therefore no exclusions were made for poor RNA quality. cDNA libraries were prepared from 5 ng total RNA with the SMART-seq2 protocol [5] and were tagmented with the Nextera XT kit (Illumina, USA). Library size selection was carried out with AMPure beads (Beckman-Coulter, USA) and library size distribution was measured with the High Sensitivity D5000 assay on a TapeStation 4200 System (Agilent, USA). Library concentration was determined using the HS dsDNA assay on a Qubit. Libraries were sequenced for SR 75 cycles on a NextSeq500 system

(Illumina) using High Output v2 chemistry. Base call files were converted to fastq format and demultiplexed using bcl2fastq v2.20.

Data pre-processing and RNA sequencing analysis

The 75 bp single-end reads were aligned to the human reference transcriptome hg38 from UCSC by kallisto v0.44.0 using default parameters. Data were imported into DESeq2 (v.1.10.1; [6]) using the TXimport (v1.2.0, [7]) package. DESeq2 was used for the calculation of normalized counts for each transcript using default parameters. All normalized transcripts with a maximum over all group means lower than 10 were excluded resulting in 33,032 present transcripts. Unwanted or hidden sources of variation, such as batch, sex and smoking status were removed using the sva package [8]. Briefly, the SVA package models the provided gene expression tables to identify and construct surrogate variables which adjust for technical artifacts (batches), as well as variation of unknown sources. Subsequently, the user can assess these variables and test how each of the undesired factors is corrected by the algorithm. The normalized rlog transformed expression values were adjusted according to the five surrogate variables identified by sva using the function removeBatchEffect from the limma package [9]. DE genes were defined by a p-value cut-off of 0.05 and an adjusted p-value (IHW) < 0.5 (independent hypothesis weighting). All present transcripts were used as input for principal component analysis. Pearson Correlation Coefficient Matrix (PCCM) and the top 25% most variable transcripts within the dataset were selected and visualized as heatmaps.

Gene set enrichment analysis (GSEA)

To test for functional enrichment between COPD (GOLD2 and GOLD3/4, respectively) and control patients, we performed GSEA [10] on all present genes of the dataset using the gene ontology set of biological processes. Information of gene ontology was obtained from the biological process gene set “c5.bp.v7.0.symbols.gmt”, downloaded from the Molecular Signatures Database (MSigDB). All present genes were used as background (universe). Lipid-related genes were extracted upon filtering gene ontology terms for the keywords “fat” and “lipid”. Data were FDR-corrected.

Linear support vector regression

Linear support vector regression [11] was employed to characterize the relative contribution of 28 different activation signatures derived from [12] to the control and COPD patients. Our normalized gene expression table was utilized as input mixture file and the published activation signatures in [12] were used to compute the relative activation signatures within bulk control and COPD samples (1,000 permutations). The union of the c5 module genes (Fig. 3A) was used as a fatty acid signature and the 20 most variable genes were visualized in a heatmap.

Filtering for TFs, epigenome, surfaceome and secretome

All present transcripts were filtered and sorted by their variance in the dataset. The 20 most variable genes of each category were selected and visualized in heatmaps. TF lists were extracted from [13], the epigenome gene list was derived from the literature, surface and secretome markers were extracted from the Human Protein Atlas [14, 15].

149

Construction of co-expressed network analysis – automated (CoCena²)

To elucidate similarities and differences within the gene expression patterns of the three different patient groups -control, GOLD2 and GOLD3/4 COPD- CoCena² was performed. Pearson correlation was calculated on the 6,000 most variable genes within the dataset using the R package Hmisc (v4.3-0; [16]). Data were filtered for significant (p-value < 0.05, Bonferroni correction p<0.05) and positive (r-value > 0) correlation values. A Pearson correlation coefficient cutoff of 0.88 was applied as this yields a scale-free network (R^2 = 0.89) with 5,209 nodes and 42,648 edges. The Group Fold Change (GFC) was calculated for each gene and each condition on the inverse logarithmic count data using the R package gtools (v3.8.2; [17]). In brief, the mean expression of each gene for each patient group versus the overall mean expression of the gene was calculated. Unbiased clustering was performed using the R package igraph (v1.2.4.1; [18]). Five different clustering algorithms, namely “cluster_label_prop”, “cluster_fast_greedy”, “cluster_louvain”, “cluster_infomap” and “cluster_walktrep”, were tested and “cluster_louvain” was selected as it achieves the highest modularity score. The mean GFC for each cluster and condition was visualized in the cluster-condition heatmap using the R package ComplexHeatmap (v2.0.0; [19]). Clusters with less than 35 genes are not shown. Hierarchical clustering was performed on genes and cluster modules using the Euclidean clustering distance and the complete-linkage clustering method. Network generation was performed with the R package igraph. The network information was imported to and exported from Cytoscape using the R package RCy3 (v2.6.2; [20]). In Cytoscape, the prefuse force-directed layout was applied to the network. Network visualization was performed using the R packages ggnetwork (v0.5.1,

[21]) and ggplot2 (v3.2.1; [22]).

Biological function-related bioinformatic analysis of network modules

GSEA was performed on the patient group-related modules identified by CoCena² using the R package ClusterProfiler (v3.12.0; [23]). The *compareCluster* function was used to determine significant enrichment (q-value < 0.1 using the Benjamini-Hochberg method) of hallmark gene sets of biological processes. Information of hallmark genes was obtained from the hallmark gene set “h.all.v6.1.symbols.gmt”, downloaded from the Molecular Signatures Database (MSigDB). All genes present in the network were used as background (universe). Interesting hallmark terms for each patient group were selected and the mean gene expression of all genes within the selected term was visualized in heatmaps using the R package pheatmap (v1.0.12; [24]). The gene expression values were scaled over the three groups. Hierarchical clustering was performed on the genes using the Euclidean clustering distance and the complete clustering method. Enriched genes were highlighted on the CoCena^2 network.

Lipidomics

Sorted AMs ($5x10^4$) were washed with PBS, suspended with 150 mM ammonium acetate (Merck, Germany) and transferred into a glass tube. After centrifugation at 300xg for 10 min with low brake, supernatants were discarded, and the pellets were frozen at -80°C until analysis. Extraction mix (Chloroform 1:5 methanol-containing internal standards: 210 pmol PE(31:1), 396 pmol PC(31:1), 98 pmol PS(31:1), 84 pmol PI(34:0) , 56 pmol PA(31:1), 51 pmol PG (28:0), 28 pmol CL(56:0), 39 pmol LPA (17:0),

35 pmol LPC(17:1), 38 pmol LPE (17:1), 32 pmol Cer(17:0), 99 pmol SM(17:0),55 pmol GlcCer(12:0), 14 pmol GM3 (18:0-D3), 359 pmol TG(47:1), 111 pmol CE(17:1), 64 pmol DG(31:1), 103 pmol MG(17:1), 724 pmol Chol(d6), 45 pmol Car(15:0)) was spiked to the pellets before 2 min sonication and centrifugation at 20,000xg for 2 min. Chloroform and 1% acetic acid were added and the samples were spun at 20,000xg for 2 min. The lower phase was transferred and let evaporate in the vacuum concentrator (45°C for 10 min). Spray buffer (8/5/1 2-propanol/methanol/water, 10 mM ammonium acetate) was added, the samples were sonicated for 5 min and analyzed separately by infusing them at 10 ul/min into a Thermo Q Exactive Plus spectrometer equipped with the HESI II ion source for shotgun lipidomics. MS1 spectra (res. 280,000) were recorded in 100 m/z windows from 200 – 1,200 m/z (pos.) and 200 – 1,700 m/z (neg.) followed by recording of MS/MS spectra (res. 70,000) by data independent acquisition in 1 m/z windows from 200 – 1,200 (pos.) and 200 – 1,700 (neg.) m/z.

Lipidomics analysis

Raw files were converted to .mzml format and imported into the LipidXplorer software for analysis using custom mfql files to identify sample lipids and internal standards. Raw data were filtered during import into LipidXplorer by global thresholding in MS1 and MS2 to remove small stochastic noise peaks. For identification, mass errors were limited to 6 ppm in MS1 and 0.005 Th in MS2. The average mass error over all identifications in MS1 was 1.76 ppm. For a valid identification, all species must show both an identification of the unfragmented peak in MS1 and at least one specific fragment in MS2. Consistency of analysis was checked by determination of the coefficient of

variation of the intensities of the spiked internal standards within one sample series and was found to be 11.7%. Despite the low number of cells, total identified lipids in samples were 5.5-15-fold over the background blank. Further, species pattern within lipid classes were checked for consistency with expectations. We found the very characteristic pattern of dominating saturated PC (but not PC-O) and saturated DAG species (32:0, 34:0) characteristic for the lung tissue, originating from lung surfactant secretion. PE and PI showed the expected dominating PUFA-containing PE(38:4) and PI(38:4). The pattern of CE showed a stronger signal for CE(18:1) than CE(18:2), ruling out major contamination with blood, in which CE(18:2) is about 3-fold more abundant than CE(18:1). Lipid species showing signals above the blank in min. 70% of the samples were included in the analysis. For further data processing, absolute amounts were calculated using the internal standard intensities followed by normalization on the sum of all measured lipid species per sample. %mol values were averaged for each patient group, log2-transformed and then used for fold change calculations. To find patient group-specific co-regulated lipid species, CoCena[2] was used as for the transcriptome analysis. Pearson correlation was calculated on all lipid species. Data were filtered for significant (p-value < 0.05, Bonferroni correction p<0.05) and positive (r-value > 0) correlation values leading to the construction of a correlation network with 501 nodes and 184 edges (Pearson correlation coefficient cut-off of 0.75). The GFC was calculated for each lipid species and each condition and clustering performed using the "cluster_louvain" algorithm. Mean GFCs for each cluster and condition are visualized in the cluster-condition heatmap. Hierarchical clustering was performed on lipid species and cluster modules using the Euclidean clustering distance and the complete-linkage

clustering method. Network generation was performed with the R package igraph.

Statistics

A two-tailed Welch's unpaired t test was used to analyze data from two groups. Equality of population variance was assessed with the F-test statistic for two independent groups. A non-parametric Wilcoxon test was used to perform a pairwise comparison between patient groups for all enriched macrophage activation signatures in Fig. 3F. For more than two groups, normality and homoscedasticity were first assessed using the Shapiro-Wilk and Levene tests in R (v3.6.1). A non-parametrical Kruskal-Wallis test with Dunn's multiple correction post hoc was used in Fig. E3 because the data did not follow a normal distribution. Statistical significance was inferred when $p<0.05$.

Data and code availability

The transcriptomic data are deposited at the European Genome-phenome Archive (EGA), which is hosted by the EBI and the CRG, under accession number EGAS00001004244. Flow cytometry files can be found at www.flowrepository.org under FR-FCM-Z2JL. Lipidomic data and annotation are provided in Tables E6 and E7. The code for CoCena[2] is publicly deposited on GitHub (https://github.com/UlasThomas/CoCena2).

REFERENCES

1. Global Initiative for Chronic Obstructive Lung Disease (GOLD) [Internet]. [cited 2021 Jan 24].Available from: https://goldcopd.org/wp-content/uploads/2020/11/GOLD-REPORT-2021-v1.1-25Nov20_WMV.pdf.

2. Haque RA, Usmani OS, Barnes PJ. Chronic idiopathic cough: a discrete clinical entity? *Chest* [Internet] American College of Chest Physicians; 2005 [cited 2020 Mar 30]; 127: 1710–1713Available from: http://www.ncbi.nlm.nih.gov/pubmed/15888850.

3. Kardos P, Dinh QT, Fuchs K-H, Gillissen A, Klimek L, Koehler M, Sitter H, Worth H. [Guidelines of the German Respiratory Society for Diagnosis and Treatment of Adults Suffering from Acute, Subacute and Chronic Cough]. *Pneumologie* [Internet] Georg Thieme Verlag; 2019 [cited 2020 Mar 30]; 73: 143–180Available from: http://www.ncbi.nlm.nih.gov/pubmed/30776835.

4. Meyer KC, Raghu G, Baughman RP, Brown KK, Costabel U, Du Bois RM, Drent M, Haslam PL, Kim DS, Nagai S, Rottoli P, Saltini C, Selman M, Strange C, Wood B. An official American Thoracic Society clinical practice guideline: The clinical utility of bronchoalveolar lavage cellular analysis in interstitial lung disease. Am. J. Respir. Crit. Care Med. 2012. p. 1004–1014.

5. Picelli S, Björklund ÅK, Faridani OR, Sagasser S, Winberg G, Sandberg R. Smart-seq2 for sensitive full-length transcriptome profiling in single cells. *Nat. Methods* [Internet] 2013 [cited 2017 Jun 20]; 10: 1096–1098Available from: http://www.nature.com/doifinder/10.1038/nmeth.2639.

6. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* [Internet] 2014 [cited 2018 Feb 14]; 15: 550Available from: http://www.ncbi.nlm.nih.gov/pubmed/25516281.

7. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. *F1000Research* [Internet] 2015 [cited 2018 Jul 23]; 4: 1521Available from: https://f1000research.com/articles/4-1521/v2.

8. Leek JT, Johnson WE, Parker HS, Jaffe AE, Storey JD. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* 2012; 28: 882–883.

9. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* Oxford University Press; 2015; 43: e47.

10. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, Paulovich A, Pomeroy SL, Golub TR, Lander ES, Mesirov JP. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* [Internet] 2005 [cited 2018 Nov 12]; 102: 15545–15550Available from: http://www.pnas.org/cgi/doi/10.1073/pnas.0506580102.

11. Newman AM, Liu CL, Green MR, Gentles AJ, Feng W, Xu Y, Hoang CD, Diehn M, Alizadeh AA. Robust enumeration of cell subsets from tissue expression profiles. *Nat. Methods* [Internet] 2015 [cited 2018 May 23]; 12: 453–457Available from: http://www.nature.com/articles/nmeth.3337.

12. Xue J, Schmidt SV, Sander J, Draffehn A, Krebs W, Quester I, De Nardo D, Gohel TD, Emde M, Schmidleithner L, Ganesan H, Nino-Castro A, Mallmann MR, Labzin L, Theis H, Kraut M, Beyer M, Latz E, Freeman TC, Ulas T, Schultze JL. Transcriptome-Based Network Analysis Reveals a Spectrum Model of Human Macrophage Activation. *Immunity* [Internet] 2014 [cited 2017 Jun 16]; 40: 274–288Available from: http://www.ncbi.nlm.nih.gov/pubmed/24530056.

13. Fulton DL, Sundararajan S, Badis G, Hughes TR, Wasserman WW, Roach JC, Sladek R. TFCat: the curated catalog of mouse and human transcription factors. *Genome Biol.* [Internet] 2009 [cited 2018 May 26]; 10: R29Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2009-10-3-r29.

14. The Human Protein Atlas - surfaceome [Internet]. [cited 2020 Jan 23].Available from: https://www.proteinatlas.org/search/protein_class:Predicted+membrane+proteins+AND+NOT+protein_class:Predicted+secreted+proteins.

15. The Human Protein Atas - secretome [Internet]. [cited 2020 Jan 23].Available from: https://www.proteinatlas.org/search/protein_class:Predicted+secreted+proteins+AND+NOT+protein_class:Predicted+membrane+proteins.

16. Harrell Jr FE, with contributions from Charles Dupont, many others. Hmisc: Harrell Miscellaneous [Internet]. 2019.Available from: https://cran.r-project.org/package=Hmisc.

17. Warnes GR, Bolker B, Lumley T. gtools: Various R Programming Tools [Internet]. 2018.Available from: https://cran.r-project.org/package=gtools.

18. Csardi G, Nepusz T. The igraph software package for complex network research. *InterJournal* [Internet] 2006; Complex Systems: 1695Available from: http://igraph.org.

19. Gu Z, Eils R, Schlesner M. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. *Bioinformatics* Oxford University Press; 2016; 32: 2847–2849.

20. Gustavsen JA, Pai S, Isserlin R, Demchak B, Pico AR. RCy3: Network biology using Cytoscape from within R. *F1000Research* [Internet] 2019 [cited 2020 Jan 22]; 8: 1774Available from: https://f1000research.com/articles/8-1774/v3.

21. Briatte F. ggnetwork: Geometries to Plot Networks with “ggplot2” [Internet]. 2016.Available from: https://cran.r-project.org/package=ggnetwork.

22. Wickham H. ggplot2: Elegant Graphics for Data Analysis [Internet]. Springer-Verlag New York; 2016.Available from: https://ggplot2.tidyverse.org.

23. Yu G, Wang L-G, Han Y, He Q-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *Omi. A J. Integr. Biol.* [Internet] Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA ; 2012 [cited 2019 Aug 1]; 16: 284–287Available from: http://www.liebertpub.com/doi/10.1089/omi.2011.0118.

24. Kolde R. pheatmap: Pretty Heatmaps [Internet]. 2019.Available from: https://cran.r-project.org/package=pheatmap.

SUPPLEMENTARY FIGURES

Fig. S1 – Representative multi-color flow cytometry analysis of the lymphoid compartment of a control patient (a). Absolute numbers of lymphoid immune cells in the BALF of control, GOLD2 and GOLD3/4 COPD patients calculated with traditional gating. Data are from 8-23 patients per group and are represented as mean ± SD (b). Bar plot of ILC concentration in BALF of control and COPD patients calculated with flow cytometry traditional gating (c). Bar plot of ILC1 concentration in the BALF of control and COPD patients calculated with flow cytometry traditional gating (d). Mean ± SD is shown from 16-17 patients and data were analysed with an unpaired two-tailed student t-test, * $p<0.05$, BALF, bronchoalveolar lavage fluid; ILC, innate lymphoid cell

Fig. S2 - Hierarchical clustering of the 25% most variable transcripts (a). Matrix of hierarchically clustered PCCM showing the distance from 0 to 100 (blue to red) based on all present transcripts (b). Hierarchical clustering of the patient group means of the top 20 most variable transcripts filtered by a list of TFs (c), known epigenetic regulators (d), surface (e) and secreted molecules (f), PCCM, Pearson correlation coefficient matrix; TF, transcription factor

SUPPLEMENTARY TABLES

Table 1 – Detailed demographics of control and COPD patients.

Table 2 – Detailed demographics of rejected patients.

Table 4 – Detailed information for the antibodies used in the study.

Table 5 – Surface markers that were used to identify major immune cell types in BALF from control and COPD patients.

Table 6 – Raw lipidomic data generated in the study.

Table 7 - Lipidomic data annotation table.